Iranian Journal of Fuzzy Systems

Volume 16, Number 4, (2019), pp. 65-81

University of Sistan and Baluchestan

Correlation coefficients of linguistic interval hesitant fuzzy sets and their application

Y. W. Xu[1], F. Y. Meng[2] and N. Wang[3]

[1,2,3] *Business School, Central South University, Changsha 410083,China*
[2] *School of Management Science and Engineering, Nanjing University of Information Science and Technology, Nanjing 210044,China*

xyw161611117@163.com, mengfanyongtjie@163.com, wlala0714@163.com

Abstract

To address the hesitancy, inconsistency and uncertainty of decision makers cognitions, linguistic interval hesitant fuzzy sets (LIHFSs) are efficient tools. This paper focuses on studying the application of LIHFSs. To do this, two correlation coefficients of LIHFSs are defined, which needn't consider the length of elements in LIHFSs or the arrangement of their possible interval values. To address the situation where the weights of elements in a set are different and correlative, two linguistic interval hesitant fuzzy Shapley weighted correlation coefficients are defined. Considering the situation where the weight information of features/attributes is partly known, programming models to determine the optimal fuzzy measures on them are constructed, respectively. After that, an approach to pattern recognition and multi-attribute decision making with linguistic interval hesitant fuzzy information is developed, respectively. Meanwhile, illustrative examples about medical diagnosis and selecting constructors for tunnel bidding are selected to verify the application of new approaches, and comparison with a previous method is offered.

Keywords: Decision making, pattern recognition, linguistic interval hesitant fuzzy set, correlation coefficient, distance measure.

1 Introduction

Decision making is one of the most common and important daily activities in human life. In the past half century, many researchers devoted themselves to studying the theory and application of decision making [6, 16, 17, 22, 38, 39, 60]. Bidding decision is an indispensable activity for bidder that plays an increasingly important role in the daily economic affairs. At present, there are five main methods for solving bidding problems under multi-risk factors, including the analytic hierarchy process (AHP) method, artificial neural network (ANN) method, fuzzy evaluation (FE) method, expert system (ES) method, and case-based reasoning (CBR) method. With the rapid development of human social-economic activities, it becomes more and more difficult to require decision makers (DMs) to provide exact attribute values. To address this issue, Zadeh's fuzzy set theory is a good tool [57]. With the development of fuzzy decision-making theory, many types of fuzzy sets are presented, which can be classified into two types: One type is the quantitative fuzzy sets, such as [1, 42, 57]; the other type is the qualitative fuzzy sets, including [12, 25, 30, 39, 47, 58]. The emergence of fuzzy variables provides a new way to solve engineering bidding.

Many researchers have noted that quantitative variables cannot be used in some complex decision-making problems. To deal with this problem, linguistic variables [58] are good choices that own many good characteristics. However, the use of linguistic variables has drawbacks in modeling and calculating information. Considering this issue, several computing with word (CW) models have been proposed to broaden the application of decision making with linguistic variables in the literature [34, 47, 49, 50, 57]. However, linguistic variables can only handle the exact qualitative information, which restricts the application. Xu [51] first noted this problem and introduced uncertain linguistic variables by extending the

Corresponding Author: F. Y. Meng

Received: September 2017; Revised: May 2018; Accepted: December 2018.

discrete linguistic term set to the continuous linguistic term set. Taking the advantages of Atanassov's intuitionistic fuzzy sets (IFSs) [1], interval-valued intuitionistic fuzzy sets (IVIFSs) [2] and linguistic variables [58], Wang and Li [43] gave intuitionistic linguistic sets (ILSs), Liu and Jin [21] proposed intuitionistic uncertain linguistic sets (IULSs), and Liu [20] introduced interval-valued intuitionistic uncertain linguistic sets (IVIULSs). As Torra [42] noted when a DM makes a judgment, several possible values may exist. Thus, Torra [42] introduced the concept of hesitant fuzzy sets (HFSs), which apply several possible values in [0, 1] to denote a judgement. After the original work of Torra [42], Chen et al. [6] further introduced interval-valued hesitant fuzzy sets (IVHFSs) that use several intervals in [0, 1] to express a judgment. As the above mentioned limitation, HFSs and IVHFSs can only denote the quantitative opinions of the DMs. Thus, Rodriguez et al. [39] proposed hesitant fuzzy linguistic term sets (HFLTSs), which use several linguistic variables to denote the qualitative hesitancy of DMs' recognitions. Following the original work of Rodriguez et al. [39], several methods for decision making with HFLTSs are developed [15, 16, 17, 38]. To express the quantitative and qualitative hesitancy of DMs, several types of hybrid hesitant fuzzy sets are proposed in the literature [14, 25, 27, 29, 31, 44].

Correlation coefficient [7, 10] is one of the most important indices that has been received considerable attention in many fields, such as digital image processing [46], clustering analysis [3, 5, 53], and decision making [18, 53]. To address fuzzy decision-making information, Dumitrescu [9] extended correlation to fuzzy situations and proposed the concept of fuzzy correlation, Yu [55] introduced fuzzy correlation and correlation coefficient to measure the interrelation of fuzzy sets, Chiang and Lin [7] defined fuzzy correlation coefficients by adopting the idea of conventional statistics, and Hong [13] adopted the weakest t-norm (TW)-based fuzzy arithmetic operations to research fuzzy correlation coefficient. Furthermore, several extended types of fuzzy correlation coefficients are proposed in the literature [4, 5, 10, 19, 24, 37, 52, 54]. Note that Xu and Xia's hesitant fuzzy correlation coefficients and Chen et al.'s interval-valued hesitant fuzzy correlation coefficients require the considered HFSs or IVHFSs to have the same length. In contrast to these correlation coefficients, Meng et al. [23, 24] further presented several new correlation coefficients of HFSs and IVHFSs that neither consider the length of (IVHFSs) HFSs nor the arrangement of the possible (interval) values.

Considering there is no research on correlation coefficient of LIHFSs, which restricts the application of LIHFSs. Following the works of Meng at al. [23, 24], this paper studies the correlation coefficients of LIHFSs and defines linguistic interval hesitant fuzzy correlation coefficients, which needn't consider the length of the compared LIHFSs or the arrangement of the possible interval values. The remainder of this paper is organized as follows:

In Section 2, several basic concepts related to IVHFSs, LIHFSs, fuzzy measures, 2-additive measures and the Shapley function are reviewed. In Section 3, a new distance measure for LIHFSs is defined, and several correlation coefficients of LIHFSs are proposed, including the geometric mean based correlation coefficient, the maximum based correlation coefficient, and the Shapley weighted correlation coefficients. In Section 4, a programming model for determining the optimal 2-additive measure on the feature set is built, and an approach to pattern recognition with LIHFSs is developed. Meanwhile, a pattern recognition problem about medical diagnosis is selected to show the application of the new method. In Section 5, programming models for determining the optimal 2-additive measure on the attribute set are constructed. Then, an approach to decision making with LIHFSs is provided. Meanwhile, an example about selecting constructors for tunnel bidding is offered to verify the application of the new method. The conclusions are made in the last.

2 Basic concepts

To cope with the situation where several possible values exist for a judgment, Torra [42] introduced hesitant fuzzy sets (HFSs) that use several values in [0, 1] to denote the possible membership degrees. Later, Chen et al. [6] extended HFSs to interval-valued hesitant fuzzy sets (IVHFSs) and gave the following definition:

Definition 2.1. [6] *Let $X = \{x_1, x_2, ..., x_n\}$ be a finite set. An IVHFS $\bar{A}$ on X is in terms of a function that when applied to X returns to a subset of D[0, 1], denoted by $\bar{A} = \left\{< x_i, \bar{h}_{\bar{A}}(x_i) > | x_i \in X\right\}$, where $\bar{h}_{\bar{A}}(x_i)$ is a set of all possible interval membership degrees of the element $x_i \in X$ to $\bar{A}$ with D[0, 1] being the set of all closed subintervals in [0, 1]. For convenience, Chen et al.* [6] *called $\bar{h}_{\bar{A}}(x_i)$ an interval-valued hesitant fuzzy element.*

To express the hesitancy and uncertainty of DMs' qualitative and quantitative cognitions, Meng et al. [31] introduced the concept of linguistic interval hesitant fuzzy sets (LIHFSs) as follows:

Definition 2.2. [31] *Let$X = \{x_1, x_2, ..., x_n\}$ be a finite set, and let $S = \{s_1, s_2, \cdots,$ $s_{2t+1}\}$ be the predefined linguistic term set. A LIHFS on X is a set that when applied to the linguistic terms of S it returns to a subset of S with several intervals in [0, 1], denoted by $LIH = \left\{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) | s_{\theta(i)} \in S\right\}$, where $\bar{h}(s_{\theta(i)}) = \{\bar{r}_{i_1}, \bar{r}_{i_2}, ..., \bar{r}_{i_m}\}$ is a set with m intervals in [0, 1] denoting the possible interval membership degrees of the linguistic term $s_{\theta(i)}$ to* LIH.

For example, to evaluate the quality of a mobile phone for the predefined linguistic term set $S = \{s_1:$ poor, s_2: slightly poor, s_3: average, s_4: slightly good, s_5: good$\}$, a DM may hesitate to give the interval [0.2, 0.3] or [0.4, 0.6] for slightly poor, the interval [0.2, 0.5], [0.6, 0.7] or [0.8, 0.9] for slightly good, and the interval [0.3, 0.4] for good. To deal with this situation, LIHFSs are effective tools, and the above evaluation for the quality of a mobile phone can be expressed by *LIH* (quality) = (s_2, [0.2, 0.3], [0.4, 0.6]), (s_4, [0.2, 0.5], [0.6, 0.7], [0.8, 0.9]), (s_5, [0.3, 0.4]).

For any LIHFS $LIH = \left\{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) | s_{\theta(i)} \in S\right\}$, when $\bar{h}(s_{\theta(i)}) = \{[1,1]\}$ for each linguistic variable $s_{\theta(i)}$, then it degenerates to a hesitant fuzzy linguistic term set (HFLTS) [39]. Furthermore, when each interval preferred degree in $\bar{h}(s_{\theta(i)})$ reduces to a real number for every linguistic variable $s_{\theta(i)}$, then it reduces to a linguistic hesitant fuzzy set (LHFS) [25].

Definition 2.3. [31] *Let LIH_1 be a LIHFS. Then, the complement of LIH_1 is denoted as:*

$$(LIH_1)^c = \left\{ s_{2t+1-\theta(i)}, \bigcup_{\bar{r}_i \in \bar{h}(s_{\theta(i)})} [1 - r_i^+, 1 - r_i^-] \, \middle| s_{\theta(i)} \in S \right\} \tag{1}$$

In practical decision-making problems, the independence between the weights of elements in a set is usually violated. To solve this issue, fuzzy measures [41] are efficient tools.

Definition 2.4. [41] *A fuzzy measure μ on the finite set $N = \{1, 2, ..., n\}$ is a set function μ: $P(N) \to [0,1]$ satisfying*
(i) $\mu(\emptyset) = 0, \mu(N) = 1$;
(ii) If $A, B \in P(N)$ and $A \subseteq B$, then $\mu(A) \le \mu(B)$, where $P(N)$ is the power set of N.

From the definition of fuzzy measures, one can find that they are defined on the power set, which cannot be used as the weights of elements directly. To overcome the issue, researchers adopted the following Shapley function [40]:

$$\varphi_i(\mu, N) = \sum_{S \subseteq N \setminus i} \frac{(n-s-1)!s!}{n!} \left(\mu(S \cup i) - \mu(S)\right) \qquad \forall i \in N \tag{2}$$

where μ is a fuzzy measure on $N = \{1, 2, ..., n\}$, s and n denote the cardinalities of S and N , respectively.

Formula (2) shows that the Shapley value is an expect value, which globally reflects the interactions among the weights of elements. Definition 2.4 shows that $\varphi_i(\mu, N) \ge 0$ for any $i \in N$ and $\sum_{i=1}^{n} \varphi_i(\mu, N) = 1$. Thus, $\{\varphi_i(\mu, N)\}_{i \in N}$ is a weight vector. When there is no interaction, the Shapley value reduces an additive weight, where $\varphi_i(\mu, N) = \mu(i) = \omega_i \ge 0$ for any $i \in N$ and $\sum_{i=1}^{n} \omega_i = 1$. Because fuzzy measures are defined on the power set, 2-additive measures [11] are good choices to reduce the complexity of solving fuzzy measures.

Definition 2.5. [11] *The fuzzy measure μ on $N = \{1, 2, ..., n\}$ is called a 2-additive measure, if, for any $S \subseteq N$ with $s \ge 2$, the following is true:*

$$\mu(S) = \sum_{\{i,j\} \subseteq S} \mu(i,j) - (s-2) \sum_{i \in S} \mu(i) \tag{3}$$

where s is the cardinality of S.

Theorem 2.6. [11] *μ is a 2-additive measure on $N = \{1, 2, ..., n\}$ if and only if, for all $i, j \in N$, the following conditions hold:*
(i) $\mu(i) \ge 0$ $\forall i \in N$;
(ii) $\sum_{\{i,j\} \subseteq N} \mu(i,j) - (n-2) \sum_{i \in N} \mu(i) = 1$;
(iii) $\sum_{i \subseteq S \setminus j} (\mu(i,j) - \mu(i)) \ge (s-2)\, \mu(j)$ $\forall j \in S \subseteq N$ and $s \ge 2$,
where s and n denote the cardinalities of S and N , respectively.

With respect to the 2-additive measure μ, the Shapley function can be expressed as follows [26]:

$$\varphi_i(\mu, N) = \frac{3-n}{2}\mu(i) + \frac{1}{2}\sum_{j \in N \setminus i}(\mu(i,j) - \mu(j)) \tag{4}$$

where $N = \{1, 2, ..., n\}$.

3 Correlation coefficients of LIHFSs

Correlation coefficients are effective tools that have been applied in many fields. In this section, we research correlation coefficients of LIHFSs, which neither consider the length of the compared LIHFSs nor the arrangement of the possible interval membership degrees. To do this, a distance measure for LIHFSs is defined. In contrast to Meng et al.'s Euclidean distance measure for LIHFSs [31], which cannot guarantee equal to zero if and only if LIHFSs are equal. For example, let $LIH_1 = \{(s_1, [0.4, 0.6])\}$ and $LIH_2 = \{(s_2, [0.2, 0.3])\}$. Following Meng et al.'s distance measure [31], we derive $ED(LIH_1, LIH_2) = 0$, by which we conclude that these two sets are equal. However, one can obviously see that these two sets are different. To overcome this disadvantage, we define a new distance measure for LIHFSs.

3.1 A distance measure

Let $LIH_1 = \left\{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) | s_{\theta(i)} \in S\right\}$ and $LIH_2 = \left\{\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) | s_{\theta(j)} \in S\right\}$ be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$.

The distance between intervals $\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})$ and $\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})$ is defined as follows [8]:

$$d(\bar{r}_{i_k}, \bar{r}_{j_p}) = \frac{1}{2}\left(| r_{i_k}^- - r_{j_p}^-| + |r_{i_k}^+ - r_{j_p}^+|\right) \tag{5}$$

Using formula (5), the distance between $\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})$ and $\bar{h}(s_{\theta(j)})$ is defined as follows [32]:

$$\overrightarrow{d\left(\bar{r}_{i_k}, \bar{h}(s_{\theta(j)})\right)} = \min_{\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})} \frac{1}{2}\left(|r_{i_k}^- - r_{j_p}^-| + |r_{i_k}^+ - r_{j_p}^+|\right) \tag{6}$$

Furthermore, the distance from $\bar{h}(s_{\theta(i)})$ to $\bar{h}(s_{\theta(j)})$ is defined as follows:

$$\overrightarrow{d\left(\bar{h}(s_{\theta(i)}), \bar{h}(s_{\theta(j)})\right)} = \frac{1}{m}\sum_{k=1:\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})}^{m} \min_{\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})} \frac{1}{2}\left(|r_{i_k}^- - r_{j_p}^-| + |r_{i_k}^+ - r_{j_p}^+|\right) \tag{7}$$

Let $S = \{s_1, s_2, ..., s_{2t+1}\}$ be the predefined linguistic term set, and let $s_{\theta(i)}, s_{\theta(j)} \in S$. Then, the distance measure between $s_{\theta(i)}$ and $s_{\theta(j)}$ is defined as follows [48]:

$$d\left(s_{\theta(i)}, s_{\theta(j)}\right) = \tfrac{1}{2t}|\theta(i) - \theta(j)| \tag{8}$$

Based on distance measures on interval possible membership degrees and on linguistic variables, the distance measure from an element in a LIHFS to another LIHFS is defined as follows:

Definition 3.1. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. The distance measure from $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1$ to LIH_2 is defined as follows:*

$$\overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_2\right)} = \min_{\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2} \frac{1}{2}\left(\left(\tfrac{1}{2t}|\theta(i) - \theta(j)| \vee \left(\frac{1}{m}\sum_{k=1:\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})}^{m} \min_{\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})} \frac{|r_{i_k}^- - r_{j_p}^-| + |r_{i_k}^+ - r_{j_p}^+|}{2}\right)\right)\right)$$

$$+\left(\tfrac{1}{2t}|\theta(i)-\theta(j)|\wedge\left(\frac{1}{m}\sum_{k=1:\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)})}^{m}\min_{\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)})}\frac{|r_{i_k}^{-}-r_{j_p}^{-}|+|r_{i_k}^{+}-r_{j_p}^{+}|}{2}\right)\right)\Bigg)$$
$$=\min_{\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_2}\frac{1}{2}\left(\tfrac{1}{2t}\,|\theta(i)-\theta(j)|+\left(\frac{1}{m}\sum_{k=1:\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)})}^{m}\min_{\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)})}\frac{|r_{i_k}^{-}-r_{j_p}^{-}|+|r_{i_k}^{+}-r_{j_p}^{+}|}{2}\right)\right) \tag{9}$$

Similarly, the distance measure from $\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_2$ to LIH_1 is defined as follows:

$$\overrightarrow{d\left(\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right),LIH_1\right)}=\min_{\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right)\in LIH_1}\frac{1}{2}\left(\tfrac{1}{2t}\,|\theta(j)-\theta(i)|+\left(\frac{1}{n}\sum_{p=1:\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)})}^{n}\min_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)})}\frac{|r_{j_p}^{-}-r_{i_k}^{-}|+|r_{j_p}^{+}-r_{i_k}^{+}|}{2}\right)\right) \tag{10}$$

Following Definition 3.1, the distance measure from one LIHFS to the other is defined as follows:

Definition 3.2. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. The distance measure from LIH_1 to LIH_2 is defined as follows:*

$$\overrightarrow{d\left(LIH_1,LIH_2\right)}=\frac{1}{|LIH_1|}\sum_{i=1}^{|LIH_1|}\overrightarrow{d\left(\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right),LIH_2\right)} \tag{11}$$

and the distance measure from LIH_2 to LIH_1 is defined as follows:

$$\overrightarrow{d\left(LIH_2,LIH_1\right)}=\frac{1}{|LIH_2|}\sum_{j=1}^{|LIH_2|}\overrightarrow{d\left(\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right),LIH_1\right)} \tag{12}$$

where $\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right)\in LIH_1$, $\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_2$, $|LIH_1|$ and $|LIH_2|$ denote the numbers of terms in the LIHFSs LIH_1 and LIH_2, respectively.

Example 3.3. *Let $S=\{s_1,s_2,s_3,s_4,s_5\}$ be the predefined linguistic term set, and let $LIH_1=\{(s_2,[0.2,0.3],[0.5,0.6]),(s_3,[0.1,][0.2],[0.4,0.5])\}$ and $LIH_2=\{(s_2,[0.3,0.4]),(s_3,[0.5,0.6],[0.7,0.8]),(s_4,[0.2,0.4])\}$.*

Following formulae (9), (10), (11) and (12), we have
$\overrightarrow{d\left(LIH_1,LIH_2\right)}=0.0938$, $\overrightarrow{d\left(LIH_2,LIH_1\right)}=0.1167$.

Example 3.3 shows that $\overrightarrow{d\left(LIH_1,LIH_2\right)}\neq\overrightarrow{d(LIH_2,LIH_1)}$, namely, distance measures listed in Definition 3.2 dissatisfy symmetry. To address this problem, we further introduce the following distance measure between LIHFSs:

Definition 3.4. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. The distance measure between LIH_1 and LIH_2 is defined as:*

$$d\left(LIH_1,LIH_2\right)=\frac{\overrightarrow{d\left(LIH_1,LIH_2\right)}+\overrightarrow{d(LIH_2,LIH_1)}}{2} \tag{13}$$

where $\overrightarrow{d\left(LIH_1,LIH_2\right)}$ and $\overrightarrow{d\left(LIH_2,LIH_1\right)}$ as shown in Definition 3.2.

In Example 3.3, we have $d\left(LIH_1,LIH_2\right)=0.105$ using formula (13).

To show the rationality of the distance measure shown in Definition 3.4, we consider the following property:

Property 3.5. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. Then,*
(i) $d(LIH_1,LIH_2)=0$ if and only if $LIH_1=LIH_2$, namely, if and only if $\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_1$ and $\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right)\in LIH_2$ for all $i=1,...,|LIH_1|$ and all $j=1,...,|LIH_2|$;
(ii) $0\leq d(LIH_1,LIH_2)\leq 1$;
(iii) $d(LIH_1,LIH_2)=d(LIH_2,LIH_1)$.

Proof. For (i): When $d(LIH_1, LIH_2) = 0$, we have $\overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_2\right)} = 0$ for any $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1$ and $\overrightarrow{d\left(\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right), LIH_1\right)} = 0$ for any $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$. Thus,

$$\min_{\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2} \frac{1}{2}\left(\tfrac{1}{2t}|\theta(i)-\theta(j)| + \left(\frac{1}{m}\sum_{k=1:\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})}^{m} \min_{\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})} \frac{|r_{i_k}^- - r_{j_p}^-| + |r_{i_k}^+ - r_{j_p}^+|}{2}\right)\right) = 0 \tag{14}$$

and

$$\min_{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1} \frac{1}{2}\left(\tfrac{1}{2t}|\theta(j)-\theta(i)| + \left(\frac{1}{n}\sum_{p=1:\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)})}^{n} \min_{\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})} \frac{|r_{j_p}^- - r_{i_k}^-| + |r_{j_p}^+ - r_{i_k}^+|}{2}\right)\right) = 0 \tag{15}$$

Following formula (14), we get $|\theta(i)-\theta(j)| = 0$ and $|r_{i_k}^- - r_{j_p}^-| = |r_{i_k}^+ - r_{j_p}^+| = 0$ for some $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$ with $\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})$, $k = 1, 2, ..., m$. Thus, $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) = \left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$.

Similarly, we obtain $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_1$ based on formula (15).

On the basis of the above discussions, we have $d(LIH_1, LIH_2) = 0$. On the other hand, from $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_1$ and $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_2$ for all $i = 1, ..., |LIH_1|$ and all $j = 1, ..., |LIH_2|$, one can easily obtain that $d(LIH_1, LIH_2) = 0$ based on formula (13).

The conditions (ii) and (iii) are obvious. □

Property 3.6. *Let LIH_1,LIH_2 and LIH_3 be any three LHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. If we have*

$$\overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_3\right)} \le \overrightarrow{d\left(\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right), LIH_3\right)} \text{ and } \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_1\right)} \le \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_2\right)}$$

for any $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1$, any $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$ and any $\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right) \in LIH_3$, then $d(LIH_1, LIH_3) \le d(LIH_2, LIH_3)$.

Proof. From Definitions 3.2 and 3.4, we obtain

$$d(LIH_1, LIH_3) = \frac{1}{2}\left(\frac{1}{|LIH_1|}\sum_{i=1}^{|LIH_1|} \overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_3\right)} + \frac{1}{|LIH_3|}\sum_{i=1}^{|LIH_3|} \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_1\right)}\right)$$

and

$$d(LIH_2, LIH_3) = \frac{1}{2}\left(\frac{1}{|LIH_2|}\sum_{i=1}^{|LIH_2|} \overrightarrow{d\left(\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right), LIH_3\right)} + \frac{1}{|LIH_3|}\sum_{i=1}^{|LIH_3|} \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_2\right)}\right)$$

Because $\overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_3\right)} \le \overrightarrow{d\left(\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right), LIH_3\right)}$ for any $\left(s_{\theta(i)}\ ,\ \bar{h}(s_{\theta(i)})\right) \in LIH_1$, and any $(s_{\theta(j)}, \bar{h}(s_{\theta(j)})) \in LIH_2$, we derive

$$\frac{1}{|LIH_1|}\sum_{i=1}^{|LIH_1|} \overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_3\right)} \le \frac{1}{|LIH_2|}\sum_{i=1}^{|LIH_2|} \overrightarrow{d\left(\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right), LIH_3\right)}$$

Similarly, we have

$$\frac{1}{|LIH_3|}\sum_{i=1}^{|LIH_3|} \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_1\right)} \le \frac{1}{|LIH_3|}\sum_{i=1}^{|LIH_3|} \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_2\right)}$$

from $\overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_1\right)} \le \overrightarrow{d\left(\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right), LIH_2\right)}$ for any $\left(s_{\theta(l)}, \bar{h}(s_{\theta(l)})\right) \in LIH_3$. Thus, $d(LIH_1, LIH_3) \le d(LIH_2, LIH_3)$. □

3.2 Correlation of LIHFSs

This subsection focuses on the correlation of LIHFSs. To do this, we first introduce the following notations:

Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. If

$$\overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), \left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right)\right)} = \overrightarrow{d\left(\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right), LIH_2\right)} \tag{16}$$

for $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1$ and $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$. Then, we denote $\left(s_{\theta(j)},\ \bar{h}(s_{\theta(j)})\right)$ as $\left(s^i_{\theta(j)}, \bar{h}^i(s_{\theta(j)})\right)$.

Formula (9) shows that $\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2$ is the element with the smallest distance to $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1$. To denote this relationship, we give the above notation.

Similarly, with respect to $\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right)$ and $\left(s^i_{\theta(j)}, \bar{h}^i(s_{\theta(j)})\right)$, if we have

$$d(\bar{r}_{i_k}, \bar{r}_{j_p}) = d\left(\bar{r}_{i_k}, \bar{h}^i(s_{\theta(j)})\right) \tag{17}$$

with $\bar{r}_{j_p} \in \bar{h}^i(s_{\theta(j)})$, then $\bar{r}_{j_p}$ is denoted by $\bar{r}^{i_k}_{j_p}$, where $\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})$, k =1, 2, ..., m.

Using these notations, the correlation of LIHFSs is defined as follows:

Definition 3.7. *Let $LIH_1 = \left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right)$ and $LIH_2 = \left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right)$ be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. Then, their correlation is defined as:*

$$\begin{aligned} C(LIH_1, LIH_2) = & \sum_{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1} \frac{1}{(2t+1)^2 \# \bar{h}(s_{\theta(i)})} \sum_{\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)}); \bar{r}^{i_k}_{j_p} \in \bar{h}^i(s_{\theta(j)})} \left(\theta(i) r^-_{i_k} \theta(j) r^{i_k,-}_{j_p} + \theta(i) r^+_{i_k} \theta(j) r^{i_k,+}_{j_p}\right) \\ & + \sum_{\left(s_{\theta(j)}, \bar{h}(s_{\theta(j)})\right) \in LIH_2} \frac{1}{(2t+1)^2 \# \bar{h}(s_{\theta(j)})} \sum_{\bar{r}_{j_p} \in \bar{h}(s_{\theta(j)}); \bar{r}^{j_p}_{i_k} \in \bar{h}^j(s_{\theta(i)})} \left(\theta(j) r^-_{j_p} \theta(i) r^{j_p,-}_{i_k} + \theta(j) r^+_{j_p} \theta(i) r^{j_p,+}_{i_k}\right) \end{aligned} \tag{18}$$

where $\#\bar{h}(s_{\theta(i)})$ and $\#\bar{h}(s_{\theta(j)})$ denote the number of elements in $\bar{h}(s_{\theta(i)})$ and $\bar{h}(s_{\theta(j)})$, respectively, and the other notations as shown above.

Example 3.8. *Let $S = \{s_1, s_2, s_3, s_4, s_5\}$ be the predefined linguistic term set. Let $LIH_1 = \{(s_1, [0.4, 05], [0.7, 0.8]),$ $(s_3, [0.2, 0.3],\ [0.5, 0.7], [0.4, 0.6])\}$ and $LIH_2 = \{(s_2, [0.2, 0.4]), (s_3, [0.5, 0.6], [0.7, 0.9])\}$ be two LIHFSs for S.*

Following formula (18), we have $C(LIH_1, LIH_2) = 0.828$.

Property 3.9. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. The correlation defined in formula (18) satisfies:*

(i) $C(LIH_1, LIH_1) = 2E(LIH_1)$ with $E(LIH_1) = \sum_{\left(s_{\theta(i)}, \bar{h}(s_{\theta(i)})\right) \in LIH_1} \frac{1}{(2t+1)^2 \# \bar{h}(s_{\theta(i)})} \sum_{\bar{r}_{i_k} \in \bar{h}(s_{\theta(i)})} \left((\theta(i))^2 (r^-_{i_k})^2 + (\theta(i))^2 (r^+_{i_k})^2\right)$;

(ii) $C(LIH_1, LIH_2) = C(LIH_2, LIH_1)$.

Proof. Following formula (18), one can easily derive the conclusions. □

3.3 Two correlation coefficients of LIHFSs

On the basis of the correlation listed in subsection 3.2, this subsection focuses on correlation coefficients of LIHFSs.

Definition 3.10. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. Then, their correlation coefficients are defined as follows:*

(i) The geometric mean based correlation coefficient:

$$CC_1(LIH_1, LIH_2) = \frac{C(LIH_1, LIH_2)}{\sqrt{E(LIH_1) E({LIH_2}^{LIH_1})} + \sqrt{E(LIH_2) E({LIH_1}^{LIH_2})}} \tag{19}$$

(ii) The maximum based correlation coefficient:

$$CC_2(LIH_1,LIH_2)=\frac{C(LIH_1,LIH_2)}{\max\left\{E(LIH_1),E({LIH_2}^{LIH_1})\right\}+\max\left\{E(LIH_2),E({LIH_1}^{LIH_2})\right\}} \tag{20}$$

where

$$E(LIH_2)=\sum_{\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_2}\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(j)})}\sum_{\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)})}\left((\theta(j))^2(r_{j_p}^-)^2+(\theta(j))^2(r_{j_p}^+)^2\right)$$

$$E({LIH_2}^{LIH_1})=\sum_{\left(s_{\theta(j)},\bar{h}(s_{\theta(j)})\right)\in LIH_2}\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(i)})}\sum_{\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(j)});\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)})}\left((\theta(j))^2(r_{j_p}^{i_k,-})^2+(\theta(j))^2(r_{j_p}^{i_k,+})^2\right)$$

$$E({LIH_1}^{LIH_2})=\sum_{\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right)\in LIH_1}\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(j)})}\sum_{\bar{r}_{i_k}^{j_p}\in\bar{h}^j(s_{\theta(i)});\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)})}\left((\theta(i))^2(r_{i_k}^{j_p,-})^2+(\theta(i))^2(r_{i_k}^{j_p,+})^2\right)$$

and the other notations as shown above.

Correlation coefficients defined in formulae (19) and (20) neither consider the length of the LIHFSs nor arrange the interval possible values in an increasing order.

To show the rationality of correlation coefficients listed in Definition 3.7, we consider the following several desirable properties:

Property 3.11. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. Then,correlation coefficients defined in formulae (19) and (20) satisfy:*
(i) $CC_1(LIH_1,LIH_1)=CC_2(LIH_1,LIH_1)=1$;
(ii) $CC_q(LIH_1,LIH_2)=CC_q(LIH_2,LIH_1)$, q = 1, 2;
(iii) $0\le CC_q(LIH_1,LIH_2)\le 1$, q = 1, 2.

Proof. The conditions (i) and (ii) are obvious.

For (iii): Following formula (19), one can find that $CC_1(LIH_1,LIH_2)\ge 0$. According to the Cauchy-Schwarz inequality

$$(x_1y_1+x_2y_2+...+x_ny_n)^2\le({x_1}^2+{x_2}^2+...+{x_n}^2)({y_1}^2+{y_2}^2+...+{y_n}^2)$$

where $x_i,y_i\in R$, $i=1,2,...,n$, we obtain

$$\left(\sum_{\left(s_{\theta(i)},\bar{h}(s_{\theta(i)})\right)\in LIH_1}\tfrac{1}{(2t+1)^2\#\bar{h}(s_{\theta(i)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(j)})}(\theta(i)r_{i_k}^-\,\theta(j)r_{j_p}^{i_k,-}+\theta(i)r_{i_k}^+\,\theta(j)\,r_{j_p}^{i_k,+})\right)^2$$

$$=\Bigg(\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}(\theta(1)r_{i_k}^-\theta(1)r_{j_p}^{i_k,-})+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}(\theta(1)r_{i_k}^+\theta(1)r_{j_p}^{i_k,+})$$

$$+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(n)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}(\theta(n)r_{i_k}^-\theta(n)r_{j_p}^{i_k,-})+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(n)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}(\theta(n)r_{i_k}^+\theta(n)r_{j_p}^{i_k,+}))\Bigg)^2$$

$$=\Bigg(\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}\frac{\theta(1)r_{j_p}^{i_k,-}}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(1)})}}\frac{\theta(1)r_{i_k}^-}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(1)})}}+\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}\frac{\theta(1)r_{i_k}^+}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(1)})}}\frac{\theta(1)r_{j_p}^{i_k,+}}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(1)})}}$$

$$+\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(n)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}\frac{\theta(n)r_{i_k}^-}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(n)})}}\frac{\theta(n)r_{j_p}^{i_k,-}}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(n)})}}+\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(n)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}\frac{\theta(n)r_{i_k}^+}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(n)})}}\frac{\theta(n)r_{j_p}^{i_k,+}}{\sqrt{(2t+1)^2\#\bar{h}(s_{\theta(n)})}}\Bigg)^2$$

$$\le\Bigg(\tfrac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)})}(\theta(1)r_{i_k}^-)^2+\tfrac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)})}(\theta(1)r_{i_k}^+)^2$$

$$+\tfrac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(1)})}(\theta(n)r_{i_k}^-)^2+\tfrac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(n)})}(\theta(n)r_{i_k}^+)^2\Bigg)\text{''}.$$

$$+\left(\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}(\theta(1)r_{j_p}^{i_k,-})^2+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(1)})}\sum_{\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(1)})}(\theta(1)r_{j_p}^{i_k,+})^2\right.$$
$$+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}(\theta(n)r_{j_p}^{i_k,-})^2$$
$$\left.+\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(n)})}\sum_{\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(n)})}(\theta(n)r_{j_p}^{i_k,+})^2\right)=E(LIH_1)E(LIH_2{}^{LIH_1})$$

Thus,

$$\sum_{(s_{\theta(i)},\bar{h}(s_{\theta(i)}))\in LIH_1}\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(i)})}\sum_{\bar{r}_{i_k}\in\bar{h}(s_{\theta(i)});\bar{r}_{j_p}^{i_k}\in\bar{h}^i(s_{\theta(j)})}(\theta(i)r_{i_k}^{-}\theta(j)r_{j_p}^{i_k,-}+\theta(i)r_{i_k}^{+}\theta(j)r_{j_p}^{i_k,+})\le\sqrt{E(LIH_1)E(LIH_2{}^{LIH_1})} \tag{21}$$

Similarly, we have

$$\sum_{(s_{\theta(j)},\bar{h}(s_{\theta(j)}))\in LIH_2}\frac{1}{(2t+1)^2\#\bar{h}(s_{\theta(j)})}\sum_{\bar{r}_{j_p}\in\bar{h}(s_{\theta(j)});\bar{r}_{i_k}^{j_p}\in\bar{h}^j(s_{\theta(i)})}(\theta(j)r_{j_p}^{-}\theta(i)r_{i_k}^{j_p,-}+\theta(j)r_{j_p}^{+}\theta(i)r_{i_k}^{j_p,+})\le\sqrt{E(LIH_2)E(LIH_1{}^{LIH_2})} \tag{22}$$

On the basis of formulae (21) and (22), we obtain

$$C(LIH_1,LIH_2)\le\sqrt{E(LIH_1)E(LIH_2{}^{LIH_1})}+\sqrt{E(LIH_2)E(LIH_1{}^{LIH_2})}$$

by which, $CC_1(LIH_1,LIH_2)\le 1$.

Similar to the proof of formula (19), one can derive the conclusions for formula (20). □

Property 3.12. *Let LIH_1 and LIH_2 be any two LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. Then, $CC_1(LIH_1,LIH_2)\ge CC_2(LIH_1,LIH_2)$.*

Formula (19) is the ratio between $C(LIH_1,LIH_2)$ and $\left(\sqrt{E(LIH_1)E(LIH_2{}^{LIH1})}+\sqrt{E(LIH_2)E(LIH_1{}^{LIH2})}\right)$, and formula (20) is the ratio between $C\left(LIH_1,LIH_2\right)$ and $\left(\max\left\{E(LIH_1),E(LIH_2{}^{LIH_1})\right\}+\max\left\{E(LIH_2),\ E(LIH_1{}^{LIH_2})\right\}\right)$. It is noteworthy that the optimistic DMs can apply the geometric mean based correlation coefficient, while the pessimistic DMs may employ the maximum based correlation coefficient.

Example 3.13. *Let LIH_1 and LIH_2 be two LIHFSs for $S=\{s_1,s_2,s_3,s_4,s_5\}$, where*

$$LIH_1=\{(s_2,[0.1,0.3],[0.5,0.7]),(s_3,[0.3,0.4],[0.6,0.8])\}$$

$$LIH_2=\{(s_1,[0.1,0.2],[0.5,0.6],[0.7,0.9]),(s_4,[0.4,0.6])\}$$

Following formula (19), the geometric mean based correlation coefficient of LIH_1 and LIH_2 is $CC_1(LIH_1,LIH_2)=0.811$;

On the basis of formula (20), the maximum based correlation coefficient of LIHFSs LIH_1 and LIH_2 is $CC_2(LIH_1,LIH_2)=0.720$.

3.4 Two shapley weighted correlation coefficients of LIHFSs

As many researchers noted, interdependences between the weights of elements in a set may exist [26, 33]. To address this issue, the Shapley function with respect to fuzzy measures is an ideal tool that has been widely used in decision making. This subsection defines two Shapley weighted correlation coefficients of LIHFSs.

Definition 3.14. *Let $LIH^1=\left\{LIH_1^1,LIH_2^1,...,LIH_n^1\right\}$ and $LIH^2=\left\{LIH_1^2,LIH_2^2,\cdots,LIH_n^2\right\}$ be two collections of LIHFSs for the predefined linguistic term set $S=\{s_1,s_2,...,s_{2t+1}\}$. The Shapley weighted correlation coefficients are defined as follows:*
(i) The geometric mean based Shapley weighted correlation coefficient:

$$CC_1^{\varphi}(LIH^1,LIH^2)=\sum\nolimits_{l=1}^{n}\varphi_l(\mu,N)CC_1(LIH_l^1,LIH_l^2)$$

(23)

(ii) The maximum based Shapley weighted correlation coefficient:

$$CC_2^{\varphi}(LIH^1, LIH^2) = \sum\nolimits_{l=1}^{n} \varphi_l(\mu, N) CC_2(LIH_l^1, LIH_l^2) \tag{24}$$

where $\varphi_l(\mu, N)$ is the Shapley value of the index l with respect to the fuzzy measure on the set $N = \{1, 2, ..., n\}$, l = 1, 2, ..., n.

Note that the Shapley weighted correlation coefficients offered in Definition 3.9 have the similar properties for linguistic interval hesitant fuzzy correlation coefficients listed in Definition 3.7.

Property 3.15. *Let $LIH^1 = \{LIH_1^1, LIH_2^1, ..., LIH_n^1\}$ and $LIH^2 = \{LIH_1^2, LIH_2^2, \cdots, LIH_n^2\}$ be two collections of LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. The Shapley weighted correlation coefficients defined in formulae (23) and (24) satisfy:*
(i) $CC_1^{\varphi}(LIH^1, LIH^1) = CC_2^{\varphi}(LIH^1, LIH^1) = 1$;
(ii) $CC_q^{\varphi}(LIH^1, LIH^2) = CC_q^{\varphi}(LIH^2, LIH^1)$, q = 1, 2;
(iii) $0 \le CC_q^{\varphi}(LIH^1, LIH^2) \le 1$, q = 1, 2.

Subsection 3.3 offers two correlation coefficients of LIHFSs based on the assumption that the weights of elements are independent and have the same importance. On the other hand, the Shapley weighted correlation coefficients (23) and (24) define the interactive weights of elements in a set. When there are no interactions between the weights of elements, the Shapley weighted correlation coefficients degenerate to the associated weighted correlation coefficients as follows:
(i) The geometric mean based weighted correlation coefficient:

$$CC_1^{\omega}(LIH^1, LIH^2) = \sum\nolimits_{l=1}^{n} \omega_l CC_1(LIH_l^1, LIH_l^2) \tag{25}$$

(ii) The maximum based weighted correlation coefficient:

$$CC_2^{\omega}(LIH^1, LIH^2) = \sum\nolimits_{l=1}^{n} \omega_l CC_2(LIH_l^1, LIH_l^2) \tag{26}$$

where ω_l is the weight of the index l, l = 1, 2 , ..., n.

Similar to the property above mentioned, we have the following properties for the (Shapley) weighted correlation coefficients.

Property 3.16. *Let $LIH^1 = \{LIH_1^1, LIH_2^1, ..., LIH_n^1\}$ and $LIH^2 = \{LIH_1^2, LIH_2^2, ..., LIH_n^2\}$ be two collections of LIHFSs for the predefined linguistic term set $S = \{s_1, s_2, ..., s_{2t+1}\}$. Then,*
(i) $CC_1^{\varphi}(LIH^1, LIH^2) \ge CC_2^{\varphi}(LIH^1, LIH^2)$;
(ii) $CC_1^{\omega}(LIH^1, LIH^2) \ge CC_2^{\omega}(LIH^1, LIH^2)$.

Similar to the analysis of correlation coefficients defined in Subsection 3.3, the pessimistic DMs can choose the geometric mean based (Shapley) weighted correlation coefficient, while the optimistic DMs prefer to adopt the maximum based (Shapley) weighted correlation coefficient.

4 An approach to pattern recognition

Pattern recognition originated from 1920s is the primary implementation approach, which aims to identify sample based on mathematical methods. This section applies the defined correlation coefficients of LIHFSs to present an approach to pattern recognition with linguistic interval hesitant fuzzy information.

Considering a pattern recognition problem, let $A = \{a_1, a_2, ..., a_m\}$ be the set of patterns, and let $C = \{c_1, c_2, ..., c_n\}$ be the set of features. A LIHFS LIH_{ij} is the evaluation of the pattern a_i with respect to the feature c_j, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$. Let $LIH_i = \{LIH_{i1}, LIH_{i2}, ..., LIH_{in}\}$, $i = 1, 2, ..., m$, be the sets of LIHFSs for patterns, and let $LIH = \{LIH_1, LIH_2, ..., LIH_n\}$ be the set of LIHFSs for the sample p with respect to the feature set C.

When the weight information is exactly known, we can apply correlation coefficients of LIHFSs offered in Section 3 to discriminate the sample p belonging to which type of patterns. Otherwise, we first need to determine the weight vector on the feature set. Since all patterns are non-inferior, and the weight vector makes the comprehensive values the bigger the better, the following programming model for the optimal fuzzy measure μ on C is established:

$$\min \sum_{j=1}^{n} \sum_{i=1}^{m} \varphi_{c_j}(\mu, C) d(LIH_{ij}, LIH_j)$$

$$s.t.\begin{cases}\mu(C)=1\\ \mu(S)\le\mu(T),\quad \forall S,T\subseteq C\ \ s.t. S\subseteq T\\ \mu(c_j)\in W_j,\mu(c_j)\ge 0,\quad j=1,2,...,n\end{cases} \tag{27}$$

where $\varphi_{c_j}(\mu,C)$ is the Shapley value of the feature c_j with respect to the fuzzy measure μ on C, W_j is the range of the known weight information.

When μ is a 2-additive measure, model (27) can be equivalently transformed into the following programming model:

$$\min\frac{3-n}{2}\sum_{j=1}^{n}\sum_{i=1}^{m}d(LIH_{ij},LIH_j)\mu(c_j)+\frac{1}{2}\sum_{j=1}^{n}\sum_{i=1}^{m}d(LIH_{ij},LIH_j)\left(\mu(c_j,c_i)-\mu(c_i)\right)$$

$$s.t.\begin{cases}\sum\limits_{c_i\subseteq S\backslash c_j}(\mu(c_j,c_i)-\mu(c_i))\ge(s-2)\mu(c_j),\quad \forall c_j\in S\subseteq C,s\ge 2\\ \sum\limits_{\{c_j,c_i\}\subseteq C}\mu(c_j,c_i)-(c-2)\sum\limits_{c_j\in C}\mu(c_j)=1\\ \mu(c_j)\in W_j,\mu(c_j)\ge 0,\quad j=1,2,...,n\end{cases} \tag{28}$$

When there is no interaction between the weights of features, models (27) and (28) reduce to programming models for the optimal additive weight vector on the feature set C.

Based on the defined correlation coefficients of LIHFSs and the built models for the optimal weight vector, the following algorithm for pattern recognition with linguistic interval hesitant fuzzy information is offered:

Step 1: Let $LIH_i=\{LIH_{i1},LIH_{i2},...,LIH_{in}\}$, $i=1,2,...,m$, be the sets of LIHFSs for the patterns in $A=\{a_1,a_2,...,a_m\}$ with respect to the feature set $C=\{c_1,c_2,...,c_n\}$, and let $LIH=\{LIH_1,LIH_2,...,LIH_n\}$ be the set of LIHFSs for the sample p with respect to the feature set C. When the weight information is partly known, model (28) is applied to determine the optimal 2-addtive measure on the feature set C. Otherwise, go to step 3;

Step 2: The Shapley weighted correlation coefficients is adopted to calculate correlation coefficient between each pattern and the sample p;

Step 3: On the basis of correlation coefficients obtained in Step 3, the sample p is recognized.

Step 4: End.

Example 4.1. *[59] We consider a set of diagnoses $A=\{a_1$: Viral fever, a_2: Malaria, a_3: Typhoid$\}$ and a set of symptoms $C=\{c_1$: Temperature, c_2 : Headache, c_3 : Cough$\}$ for the predefined linguistic term set $S=\{s_1$: poor, s_2: slightly poor, s_3: fair, s_4: slightly good, s_5: good$\}$.*

Suppose that a patient p with respect to the features in C is a set of LIHFSs, where
LIH *(patient)* $=\{\langle c_1,\{(s_2,[0.1,0.2],[0.3,0.5])\}\rangle,\langle c_2,\{(s_1,[0,6,0.8]),$
$(s_4,[0.2,0.3],[0.4,0.5])\rangle,\langle c_3,\{(s_5,[0.4,0.6])\}\rangle\}$.

The LIHFSs for each diagnosis a_i , i = 1, 2, 3, with respect to the features in C are offered as follows:

$$LIH_1(\text{Viral fever})=\{\langle c_1,\{(s_2,[0.3,0.5],[0.6,0.8])\}\rangle,$$
$$\langle c_2,\{(s_1,[0.2,0.4],[0.4,0.5]),(s_2,[0.3,0.5])\}\rangle,\langle c_3,\{(s_4,[0.4,0.5])\}\rangle\}$$
$$LIH_2(\text{Malaria})=\{\langle c_1,\{(s_5,[0.1,0.3],[0.4,0.5])\}\rangle,$$
$$\langle c_2,\{(s_2,[0.3,0.6],[0.7,0.8]),(s_3,[0.7,0.9])\}\rangle,\langle c_3,\{(s_1,[0.8,0.9])\}\rangle\}$$
$$LIH_3(\text{Typhoid})=\{\langle c_1,\{(s_2,[0.6,0.7],[0.8,0.9])\}\rangle,$$
$$\langle c_2,\{(s_2,[0.1,0.2],[0.3,0.5]),(s_5,[0.5,0.6])\}\rangle,\langle c_3,\{(s_3,[0.5,0.7])\}\rangle\}$$

Furthermore, the weight vector on the feature set C is defined as W_c = ([0.1, 0.4], [0.3, 0.4], [0.25, 0.3]).

To recognize the patient p belonging to which diagnosis, the following procedure is offered:

Step 1: Following model (28), we get:
$\mu(c_1)=0.4$, $\mu(c_2)=0.3$, $\mu(c_3)=0.25$, $\mu(c_1,c_2)=0.7$, $\mu(c_1,c_3)=0.7$, $\mu(c_2,c_3)=0.55$.

Step 2: The Shapley values using formula (2) are derived as follows:
$\varphi_{c_1}(\mu,C)=0.25$, $\varphi_{c_2}(\mu,C)=0.4$, $\varphi_{c_3}(\mu,C)=0.35$.

Meanwhile, the geometric mean based Shapley weighted correlation coefficients are
$CC_1^{\varphi}(LIH_1,LIH)=0.937$, $CC_1^{\varphi}(LIH_2,LIH)=0.865$, $CC_1^{\varphi}(LIH_3,LIH)=0.974$.
and the maximum based Shapley weighted correlation coefficients are
$CC_2^{\varphi}(LIH_1,LIH)=0.618$, $CC_2^{\varphi}(LIH_2,LIH)=0.452$, $CC_2^{\varphi}(LIH_3,LIH)=0.615$.

Step 3: On the basis of $CC_1^{\varphi}(LIH_i,LIH)$, i = 1, 2, 3, we conclude that the patient p had typhoid. However, the patient p had viral fever based on $CC_2^{\varphi}(LIH_i,LIH)$, i = 1, 2, 3.

Step 4: End.

The result shows that different optimal patterns may be derived for different correlation coefficients. Thus, the DMs need to choose the used correlation coefficients following their risk preferences and decision-making problems.

5 An approach to decision making with LIHFSs

This section considers decision making with linguistic interval hesitant fuzzy information. Without loss of generality, let $S = \{s_1, s_2, ..., s_{2t+1}\}$ be the predefined linguistic term set, let $A = \{a_1, a_2, ..., a_m\}$ be the set of alternatives, and let $C = \{c_1, c_2, ..., c_n\}$ be the set of attributes. Suppose that LIH_{ij} is the LIHFS of the alternative a_i for the attribute c_j, $i = 1, 2, ..., m$; $j = 1, 2, ..., n$. By $LIH = (LIH_{ij})_{m\times n}$, we denote linguistic interval hesitant fuzzy decision matrix.

5.1 Models to determine the optimal weight vector

As the analysis for pattern recognition, when the weight information is incompletely known, we first need to determine the weight vector on the attribute set.

Let LIH_j be the jth column of the linguistic interval hesitant fuzzy decision matrix LIH, and let

$$LIH_j^+ = \bigcup_{i=1}^m LIH_{ij} = \{(s_{\max_{i=1}^m \theta(ij)}, [\max_{i=1}^m\{r_{ij}^-\}, \quad \max\{r_{ij}^+\}])|\bar{r}_{ij} = [r_{ij}^-, r_{ij}^+] \in LIH_{ij}, i = 1, 2, ..., m\}$$

$$LIH_j^- = \bigcup_{i=1}^m LIH_{ij} = \{(s_{\min_{i=1}^m \theta(ij)}, [\min_{i=1}^m\{r_{ij}^-\}, \quad \min_{i=1}^m\{r_{ij}^+\}])|\bar{r}_{ij} = [r_{ij}^-, r_{ij}^+] \in LIH_{ij}, i = 1, 2, ..., m\}$$

for each $j = 1, 2, ..., n$ and $S_{\theta(ij)} \in S$.

Let $LIH^+ = \{LIH_1^+, LIH_2^+, ..., LIH_n^+\}$ and $LIH^- = \{LIH_1^-, LIH_2^-, ..., LIH_n^-\}$.

When the weight information on the attribute set is incompletely known, we establish the following programming model for obtaining the optimal fuzzy measure μ on C:

$$\min \sum_{j=1}^n \sum_{i=1}^m \varphi_{c_j}(\mu, C) \frac{d(LIH_{ij}, LIH_j^+)}{d(LIH_{ij}, LIH_j^+) + d(LIH_{ij}, LIH_j^-)}$$

$$s.t. \begin{cases} \mu(C) = 1 \\ \mu(S) \le \mu(T), \quad \forall S, T \subseteq C \;\; s.t. S \subseteq T \\ \mu(c_j) \in W_j, \mu(c_j) \ge 0, \quad j = 1, 2, ..., n \end{cases} \tag{29}$$

where $\varphi_{c_j}(\mu, C)$ is the Shapley value of the attribute c_j with respect to the fuzzy measure μ on C, $d(LIH_{ij}, LIH_j^+)$ is the distance measure between LIH_{ij} and LIH_j^+, $d(LIH_{ij}, LIH_j^-)$ is the distance measure between LIH_{ij} and LIH_j^-, and W_j is the known weight information.

Similarly, we derive the following programming model for ascertaining the optimal 2-additive measure μ on C:

$$\min \frac{3-n}{2} \sum_{j=1}^n \sum_{i=1}^m \frac{d(LIH_{ij}, LIH_j^+)}{d(LIH_{ij}, LIH_j^+) + d(LIH_{ij}, LIH_j^-)} \mu(c_j)$$

$$+\frac{1}{2} \sum_{j=1}^n \sum_{i=1}^m \frac{d(LIH_{ij}, LIH_j^+)}{d(LIH_{ij}, LIH_j^+) + d(LIH_{ij}, LIH_j^-)} (\mu(c_j, c_i) - \mu(c_i))$$

$$s.t. \begin{cases} \sum\limits_{c_i \subseteq S\backslash c_j} (\mu(c_j, c_i) - \mu(c_i)) \ge (s-2)\mu(c_j), \quad \forall c_j \in S \subseteq C, s \ge 2 \\ \sum\limits_{\{c_j, c_i\} \subseteq C} \mu(c_j, c_i) - (c-2) \sum\limits_{c_j \in C} \mu(c_j) = 1 \\ \mu(c_j) \in W_j, \mu(c_j) \ge 0, \quad j = 1, 2, ..., n \end{cases} \tag{30}$$

where the notations as shown in model (29).

When there are no interactions between the weights of attributes, models (29) and (30) degenerate to models for the optimal additive measure on the attribute set C.

5.2 An algorithm

Following the above discussions, this subsection introduces the following algorithm for multi-attribute decision making with linguistic interval hesitant fuzzy information.

Step 1: Let $LIH = (LIH_{ij})_{m\times n}$ be the linguistic interval hesitant fuzzy decision matrix. When all attributes are benefit, the attribute values needn't transformation. Otherwise, we need to transform $LIH = (LIH_{ij})_{m\times n}$ into $LIH' = (LIH'_{ij})_{m\times n}$, where $LIH'_{ij} = \begin{cases} LIH_{ij} & c_j \text{ is a benefit attribute} \\ (LIH_{ij})^c & c_j \text{ is a cost attribute} \end{cases}$ for all $i = 1, 2, ..., m$; $j = 1, 2, ..., n$.

Step 2: When the weight information is exactly known, go to step 4. Otherwise, model (30) is used to determine the optimal fuzzy measure μ on the attribute set C;

Step 3: Formulae (23) or (24) and (25) or (26) are adopted to calculate the (Shapley) weighted correlation coefficients between LIH_i' and LIH^+ as well as LIH_i' and LIH^-, $i = 1, 2, ..., m$;

Step 4: The following formulae are applied to calculate the ranking values:

$$r_i^{CC_p^{\varphi}} = \frac{CC_p^{\varphi}(LIH_i', LIH^+)}{CC_p^{\varphi}(LIH_i', LIH^+) + CC_p^{\varphi}(LIH_i', LIH^-)} \qquad p = 1, 2 \tag{31}$$

$$r_i^{CC_p^{\omega}} = \frac{CC_p^{\omega}(LIH_i', LIH^+)}{CC_p^{\omega}(LIH_i', LIH^+) + CC_p^{\omega}(LIH_i', LIH^-)} \qquad p = 1, 2 \tag{32}$$

for all $i = 1, 2, ..., m$, where LIH_i' denotes the ith row of LIH';

Step 5: Following the ranking values of $r_i^{CC_p^{\varphi}}$ or $r_i^{CC_p^{\omega}}$, $i = 1, 2, ..., m$; $p = 1$, 2, the best choice is obtained;

Step 6: End.

Example 5.1. *Considering a tunnel bidding, four constructors $A = \{a_1, a_2, a_3, a_4\}$ submit their tenders for building a tunnel. To evaluate these constructors, the following four attributes are considered c_1: credit (Bidder's credit is an important index for bidding evaluation, the higher the bidder's credit rate is, the higher the winning probability will be); c_2: tender offer (The tender offer is the total price of a contract that is calculated and determined by contractor, the traditional approach adopts the principle of minimum tender offer); c_3: the limitation of project time (The limitation of project time is one of the important indexes to measure construction enterprises. The length of project time directly affects the economic benefits of construction enterprises. Under the premise of ensuring quality, the shorter the time is, the higher the economic benefits will be); c_4: the scale of company (Generally speaking, companies with large scale, good equipment, high-quality talents and good management measures have a certain influence on winning). To avoid influencing each other, the DMs are required to provide their preferences anonymously. Because of the limited expertise and the complexity of considering attributes, it is difficult or impossible to require the DMs to offer quantitative judgments. To address this issue, the DMs are allowed to apply linguistic variables. Meanwhile, the DMs may also want to express their quantitative judgments for the given linguistic variables. Furthermore, the DMs may hesitate on several possible values. To cope with these situations, LIHFSs are good tools. Let $S = \{s_1$: poor, s_2: slightly poor, s_3: fair, s_4: slightly good, s_5 : good$\}$ be the predefined linguistic term set. Suppose that the linguistic interval hesitant fuzzy decision matrix $LIH = (LIH_{ij})_{4\times 4}$ is offered in Table 1.*

	c_1	c_2	c_3	c_4
a_1	$\{(s_2, [0.1, 0.3]$, $[0.5, 0.7])$, $(s_4, [0.2, 0.3])\}$	$\{(s_3, [0.7, 0.3])\}$	$\{(s_1, [0.6, 0.7])\}$	$\{(s_5, [0.3, 0.5]$, $[0.7, 0.8])\}$
a_2	$\{(s_5, [0.4, 0.6]$, $[0.8, 0.9])\}$	$\{(s_4, [0.2, 0.5])\}$	$\{(s_3, [0.1, 0.2])\}$	$\{(s_2, [0.3, 0.5]$, $[0.5, 0.7])\}$
a_3	$\{(s_2, [0.5, 0.7])$, $(s_4, [0.2, 0.3])\}$	$\{(s_3, [0.1, 0.4])\}$	$\{(s_3, [0.2, 0.4]$, $[0.5, 0.6])\}$	$\{(s_4, [0.2, 0.4])$, $[0.7, 0.8])\}$
a_4	$\{(s_5, [0.3, 0.5])$, $[0.4, 0.6])\}$	$\{(s_4, [0.6, 0.7]$, $[0.8, 0.9])\}$	$\{(s_3, [0.8, 0.9])\}$	$\{(s_1, [0.5, 0.7])\}$

Table 1: Linguistic interval hesitant fuzzy decision matrix LIH

To select the best constructor, the following steps are needed:

Step 1: Because c_2 and c_3 are cost attributes, we need to transform the linguistic interval hesitant fuzzy decision matrix LIH into the normalized linguistic interval hesitant fuzzy decision matrix LIH' as shown in Table 2:

Step 2: Suppose that weight vector on the attribute set is defined as W_c = ([0.15, 0.2], [0.3, 0.4], [0.25, 0.3], [0.2, 0.25]). According to model (30), we obtain:

$\mu(c_1) = 0.2$, $\mu(c_2) = 0.3$, $\mu(c_3) = 0.25$, $\mu(c_4) = 0.2$, $\mu(c_1, c_2) = 0.55$, $\mu(c_1, c_3) = 0.45$, $\mu(c_2, c_3) = 0.55$, $\mu(c_1, c_4) = 0.4$, $\mu(c_2, c_4) = 0.5$, $\mu(c_3, c_4) = 0.45$.

Furthermore, the Shapley values are

$\varphi_{c_1}(\mu, C) = 0.22$, $\varphi_{c_2}(\mu, C) = 0.33$, $\varphi_{c_3}(\mu, C) = 0.25$, $\varphi_{c_4}(\mu, C) = 0.2$.

Step 3: The geometric mean based Shapley weighted correlation coefficients are

$CC_1^{\varphi}(LIH_1', LIH^+) = 0.963$, $CC_1^{\varphi}(LIH_2', LIH^+) = 0.957$, $CC_1^{\varphi}(LIH_3', LIH^+) = 0.952$, $CC_1^{\varphi}(LIH_4', LIH^+) = 0.978$, $CC_1^{\varphi}(LIH_1', LIH^-) = 0.822$, $CC_1^{\varphi}(LIH_2', LIH^-)$

	c_1	c_2	c_3	c_4
a_1	$\{(s_2,[0.1,0.3],[0.5,0.7]),(s_4,[0.2,0.3])\}$	$\{(s_2,[0.3,0.7])\}$	$\{(s_4,[0.3,0.4])\}$	$\{(s_5,[0.3,0.5],[0.7,0.8])\}$
a_2	$\{(s_5,[0.4,0.6],[0.8,0.9])\}$	$\{(s_1,[0.5,0.8])\}$	$\{(s_2,[0.8,0.9])\}$	$\{(s_2,[0.3,0.5],[0.5,0.7])\}$
a_3	$\{(s_2,[0.5,0.7]),(s_4,[0.2,0.3])\}$	$\{(s_2,[0.6,0.9])\}$	$\{(s_2,[0.6,0.8],[0.4,0.5])\}$	$\{(s_4,[0.2,0.4],[0.7,0.8])\}$
a_4	$\{(s_5,[0.3,0.5],[0.4,0.6])\}$	$\{(s_1,[0.3,0.4],[0.1,0.2])\}$	$\{(s_2,[0.1,0.2])\}$	$\{(s_1,[0.5,0.7])\}$

Table 2: Normalized linguistic interval hesitant fuzzy decision matrix LIH'

$= 0.889$, $CC_1^{\varphi}(LIH_3, LIH^-) = 0.842$, $CC_1^{\varphi}(LIH'_4, LIH^-) = 0.979$.

and the maximum based Shapley weighted correlation coefficients are

$CC_2^{\varphi}(LIH'_1, LIH^+) = 0.554$, $CC_2^{\varphi}(LIH'_2, LIH^+) = 0.461$, $CC_2^{\varphi}(LIH'_3, LIH^+) = 0.615$, $CC_2^{\varphi}(LIH'_4, LIH^+) = 0.244$, $CC_2^{\varphi}(LIH'_1, LIH^-) = 0.228$, $CC_2^{\varphi}(LIH'_2, LIH^-) = 0.210$, $CC_2^{\varphi}(LIH'_3, LIH^-) = 0.221$, $CC_2^{\varphi}(LIH'_4, LIH^-) = 0.611$.

Step 4: Using formulae (31) and (32), the following ranking values are derived: $r_1^{CC_1^{\varphi}} = 0.540$, $r_2^{CC_1^{\varphi}} = 0.518$, $r_3^{CC_1^{\varphi}} = 0.531$, $r_4^{CC_1^{\varphi}} = 0.500$, $r_1^{CC_2^{\varphi}} = 0.708$, $r_2^{CC_2^{\varphi}} = 0.686$, $r_3^{CC_2^{\varphi}} = 0.736$, $r_4^{CC_2^{\varphi}} = 0.285$.

Step 5: Following the values of $r_i^{CC_1^{\varphi}}$, *i = 1, 2, 3, 4, the constructor* a_1 *is the best choice. On the basis of the values of* $r_i^{CC_2^{\varphi}}$, *i = 1, 2, 3, 4, the constructor* a_3 *is the best choice.*

Step 6: End.

In this example, when we assume that there is no interaction between the weights of attributes, the ranking values are obtained as follows:

$r_1^{CC_1^{\omega}} = 0.540$, $r_2^{CC_1^{\omega}} = 0.518$, $r_3^{CC_1^{\omega}} = 0.541$, $r_4^{CC_1^{\omega}} = 0.500$, $r_1^{CC_2^{\omega}} = 0.717$, $r_2^{CC_2^{\omega}} = 0.682$, $r_3^{CC_2^{\omega}} = 0.745$, $r_4^{CC_2^{\omega}} = 0.276$,

by which we drive the same ranking order $r_3^{CC_p^{\omega}} \succ r_1^{CC_p^{\omega}} \succ r_2^{CC_p^{\omega}} \succ r_4^{CC_p^{\omega}}$, *p = 1, 2, namely, the constructor* a_3 *is the best choice.*

5.3 Comparison analysis

In Example 5.1, when Meng et al.'s method [31] is applied, the final ranking values are obtained as follows:

(i) The ranking values obtained from the generalized linguistic interval hesitant fuzzy hybrid weighted averaging (GLIHFHWA) operator are:

$\bar{E}(LIH_1) = [1.1, 1.85]$, $\bar{E}(LIH_2) = [1.42, 1.95]$, $\bar{E}(LIH_3) = [2.93, 4.03]$, $\bar{E}(LIH_4) = [0.61, 0.83]$;

(ii) The ranking values obtained from the generalized linguistic interval hesitant fuzzy hybrid 2-additive Shapley averaging (GLIHFH2SA) operator are:

$\bar{E}(LIH_1) = [1.09, 1.92]$, $\bar{E}(LIH_2) = [1.5, 1.88]$, $\bar{E}(LIH_3) = [1.43, 1.99]$, $\bar{E}(LIH_4) = [0.56, 0.79]$, where $\lambda = 2$.

One can see that the same ranking order $LIH_3 \succ LIH_2 \succ LIH_1 \succ LIH_4$ is obtained, namely, the constructor a_3 is the best choice.

For simplicity, ranking orders with respect to the different methods are listed in Table 3.

Table 3 shows that different methods might lead to different ranking orders and different optimal choices. When the DMs make decisions, they are required to choose decision-making method according to the actual requirement. Generally speaking, we suggest the DMs to apply the Shapley weighted correlation coefficients because they can deal with the interactive characteristics between elements in a set and seem to be simpler than Meng et al.'s method [31] in some aspects.

6 Conclusions

LIHFSs are efficient tools in dealing with qualitative and quantitative vague information that appears in the engineering and scientific applications. On the other hand, correlation coefficient is one of the widely used techniques in decision-making theory. This paper took the advantages of them and presented two correlation coefficients of LIHFSs by using the defined distance measure. Meanwhile, two linguistic interval hesitant fuzzy Shapley weighted correlation coefficients are defined to address the situation where the weights of elements in a set are interactive. It's worth noting that the Shapley weighted correlation coefficient can be seen as an extension of correlation coefficients. Considering the situation where the weight information is incompletely known, models for optimal 2-additive measures on the feature set and on the attribute set are constructed. Furthermore, a method to pattern recognition and decision making under linguistic interval hesitant fuzzy environment with incomplete weight information is developed, respectively. Meanwhile, associated practical examples are offered to show the concrete applications of the developed theoretical results and a comparison analysis is offered.

It is noteworthy that new results extend the application of LIHFSs. In the future, we will continue to research correlation coefficients of other types of fuzzy sets, such as hesitant linguistic intuitionistic fuzzy sets (HLIFSs) [44], linguistic hesitant fuzzy

Methods	**Ranking values of a_1**	**Ranking values of a_2**	**Ranking values of a_3**	**Ranking values of a_4**	**Ranking orders**
New method by using the geometric mean based Shapley weighted correlation coefficient	0.540	0.518	0.531	0.500	$r_1 \succ r_3 \succ r_2 \succ r_4$
New method by using the maximum based Shapley weighted correlation coefficient	0.708	0.686	0.736	0.285	$r_3 \succ r_1 \succ r_2 \succ r_4$
New method by using the geometric mean based weighted correlation coefficient	0.540	0.518	0.541	0.500	$r_3 \succ r_1 \succ r_2 \succ r_4$
New method by using the maximum based Shapley weighted correlation coefficient	0.717	0.682	0.745	0.276	$r_3 \succ r_1 \succ r_2 \succ r_4$
Meng et al.'s method [31] with the GLIHFHWA operator	[1.1, 1.85]	[1.42, 1.95]	[2.93, 4.03]	[0.61, 0.83]	$LIH_3 \succ LIH_2 \succ LIH_1 \succ LIH_4$
Meng et al.'s method [31] with the GLIHFH2SA operator	[1.09, 1.92]	[1.5, 1.88]	[1.43, 1.99]	[0.56, 0.79]	$LIH_3 \succ LIH_2 \succ LIH_1 \succ LIH_4$

Table 3: Ranking orders with respect to the different aggregation indices

sets (LHFSs) [25], and interval-valued hesitant fuzzy linguistic sets (IVHFLSs) [45]. Besides the theory aspect, we will study their applications in some fields, including software evaluation [5], information system selection [36], hotel selection [56] and coal mine safety evaluation [35].

Acknowledgement

This work was supported by the National Natural Science Foundation of China (Nos. 71571192, and 71671188), the Innovation-Driven Project of Central South University (No. 2018CX039), the Major Project for National Natural Science Foundation of China (No. 71790615), the Beijing Intelligent Logistics System Collaborative Innovation Center (No. 2018KF-06), the National Social Science Foundation of China (No. 16BJY119) and the Hunan Province Foundation for Distinguished Young Scholars of China (No. 2016JJ1024).

References

[1] K. T. Atanassov, *Intuitionistic fuzzy sets*, Fuzzy Sets and Systems, **20**(1) (1986), 87–96.

[2] K. T. Atanassov, G. Gargov, *Interval valued intuitionistic fuzzy sets*, Fuzzy Sets and Systems, **31**(3) (1989), 343–349.

[3] P. Bonizzoni, G. D. Vedova, R. Dondi, T. Jiang, *Correlation clustering and consensus clustering*, Proceedings of Algorithms and Computation ISAAC. Springer, Hainan, China, **38**(27) (2005), 226–235.

[4] H. Bustince, P. Burillo, *Correlation of interval-valued intuitionistic fuzzy sets*, Fuzzy Sets and Systems, **74**(2) (1995), 237–244.

[5] N. Chen, Z. S. Xu, M. M. Xia, *Correlation coefficients of hesitant fuzzy sets and their applications to clustering analysis*,Applied Mathematical Modelling, **37**(4) (2013), 2197–2211.

[6] N. Chen, Z. S. Xu, M. M. Xia, *Interval-valued hesitant preference relations and their applications to group decision making*, Knowledge-Based Systems, **37** (2013), 528–540.

[7] D. A. Chiang, N. P. Lin, *Correlation of fuzzy sets*, Fuzzy Sets and Systems, **102**(2) (1999), 221–226.

[8] Y. G. Dang, S. F. Liu, J. Song, *Study on the multi-attribute decision model of grey target based on interval numbers*, Proceedings of the IEEE Inter. Conf. Grey Syst. Intell. Ser., Nanjing, China, (2007), 744–747.

[9] D. Dumitrescu, *Fuzzy correlation*, Studia University, Babe-Bolyai Mathematica, **23** (1978), 41–44.

[10] T. Gerstenkorn, J. Mako, *Correlation of intuitionistic fuzzy sets*, Fuzzy Sets and Systems, **44**(1) (1991), 39–43.

[11] M. Grabisch, *K-order additive discrete fuzzy measures and their representation*, Fuzzy Sets and Systems, **92**(2) (1997), 167–189.

[12] F. Herrera, L. Martinez, *A 2-tuple fuzzy linguistic representation model for computing with words*, IEEE Transactions on Fuzzy Systems, **8**(6) (2000), 746–752.

[13] D. H. Hong, *Fuzzy measures for a correlation coefficient of fuzzy numbers under TW (the weakest t-norm)-based fuzzy arithmetic operations*, Information Sciences, **176**(2) (2006), 150–160.

[14] P. Ji, H. Y. Zhang, J. Q. Wang, *A projection-based outranking method with multi-hesitant fuzzy linguistic term sets for hotel location selection*, Cognitive Computation, (2018), 1–15.

[15] R. X. Liang, J. Q. Wang, H. Y. Zhang, *Projection-based PROMETHEE methods based on hesitant fuzzy linguistic term sets*, International Journal of Fuzzy Systems, **1** (2017), 1–14.

[16] H. C. Liao, Z. S. Xu, X. J. Zeng, *Distance and similarity measures for hesitant fuzzy linguistic term sets and their application in multi-criteria decision making*, Information Sciences, **271** (2014), 125–142.

[17] H. C. Liao, Z. S. Xu, X. J. Zeng, *Hesitant fuzzy linguistic VIKOR method and its application in qualitative multiple criteria decision making*, IEEE Transactions on Fuzzy Systems, **23**(5) (2015), 1343–1355

[18] H. C. Liao, Z. S. Xu, X. J. Zeng, *Novel correlation coefficients between hesitant fuzzy sets and their application in decision making*, Knowledge-Based Systems, **82** (2015), 115–127.

[19] H. C. Liao, Z. S. Xu, X. J. Zeng, J. M. Merig, *Qualitative decision making with correlation coefficients of hesitant fuzzy linguistic term sets*, Knowledge-Based Systems, **76** (2015), 127–138.

[20] P. D. Liu, *Some geometric aggregation operators based on interval intuitionistic uncertain linguistic variables and their application to group decision making*, Applied Mathematical Modelling, **37**(4) (2013), 2430–2444.

[21] P. D. Liu, F. Jin, *Methods for aggregating intuitionistic uncertain linguistic variables and their application to group decision making*, Information Sciences, **205** (2012), 58–71.

[22] F. Y. Meng, X. H. Chen, *A robust additive consistency-based method for decision making with triangular fuzzy reciprocal preference relations*, Fuzzy Optimization and Decision Making, **17** (1) (2018), 49–73.

[23] F. Y. Meng, X. H. Chen, *Correlation coefficients of hesitant fuzzy sets and their application based on fuzzy measures*, Cognitive Computation, **7**(4) (2015), 445–463.

[24] F. Y. Meng, X. H. Chen, *Correlation coefficient of interval-valued intuitionistic uncertain linguistic sets and its application*, Cybernetics and Systems, **48**(2) (2017), 114–135.

[25] F. Y. Meng, X. H. Chen, Q. Zhang, *Multi-attribute decision analysis under a linguistic hesitant fuzzy environment*, Information Sciences, **267** (2014), 287–305.

[26] F. Y. Meng, J. Tang, *Interval-valued intuitionistic fuzzy multi-attribute group decision making based on cross entropy measure and Choquet integral*, International Journal of Intelligent Systems, **28**(12) (2013), 1172–1195.

[27] F. Y. Meng, J. Tang, *New ranking order for linguistic hesitant fuzzy sets*, Journal of the Operational Research Society, **4** (2018), 1–10.

[28] F. Y. Meng, J. Tang, Q. X. An, X. H. Chen, *Decision making with intuitionistic linguistic preference relations*, International Transactions in Operational Research, **00** (2017), 1–28.

[29] F. Y. Meng, J. Tang, C. L. Li, *Uncertain linguistic hesitant fuzzy sets and their application in multi- attribute decision making*, International Journal of Intelligent Systems, **33**(3) (2018), 586–614.

[30] F. Y. Meng, J. Tang, Z. S. Xu, *Exploiting the priority weights from interval linguistic fuzzy preference relations*, Soft Computing, **2** (2017), 1–15.

[31] F. Y. Meng, C. Wang, X. H. Chen, *Linguistic interval hesitant fuzzy sets and their application in decision making*, Cognitive Computation, **8**(1) (2016), 52–68.

[32] F. Y. Meng, C. Wang, X. H. Chen, Q. Zhang, *Correlation coefficients of interval-valued hesitant fuzzy sets and their application based on the Shapley function*, International Journal of Intelligent Systems, **31**(1) (2016), 17–43.

[33] F. Y. Meng, Y. G. Yuan, X. H. Chen, *Some generalized interval-valued 2-tuple linguistic interval distance measures and their application*, International Journal of Uncertainty, Fuzziness and Knowledge-Based on systems, **25** (2017), 759–786.

[34] J. A. Morente-Molinera, J. Mezei, C. Carlsson, E. Herrera-Viedma, *Improving supervised learning classification methods using multi- granular linguistic modelling and fuzzy entropy*, IEEE Transactions on Fuzzy Systems, **25**(5) (2017), 1078–1089.

[35] H. G. Peng, J. Q. Wang, P. F. Cheng, *A linguistic intuitionistic multi-criteria decision-making method based on the Frank Heronian mean operator and its application in evaluating coal mine safety*, International Journal of Machine Learning and Cybernetics, **9**(6) (2017), 1053–1068.

[36] G. Qian, H. Wang, X. Q. Feng, *Generalized hesitant fuzzy sets and their application in decision support system*, Knowledge-Based on systems, **37** (2013), 357–365.

[37] G. H. Qu, W. H. Qu, Z. H. Zhang, et al., *Choquet integral correlation coefficient of intuitionistic fuzzy sets and its applications*, Journal of Intelligent & Fuzzy Systems, **33**(1) (2017), 1–11.

[38] R. M. Rodriǵuez, L. Martinez, F. Herrera, *A group decision making model dealing with comparative linguistic expressions based on hesitant fuzzy linguistic term sets*, Information Sciences, **241**(1) (2013), 28–42.

[39] R. M. Rodriǵuez, L. Martinez, F. Herrera, *Hesitant fuzzy linguistic term sets for decision making*, IEEE Transactions on Fuzzy Systems, **20**(1) (2012), 109–119.

[40] L. S. Shapley, *A value for n-person game*, Princeton University Press, **28** (7) (1953), 307–318.

[41] M. Sugeno, *Theory of fuzzy integral and its applications*, Ph.D. Doctorial Dissertation, Tokyo Institute of Technology, (1974), 1–123.

[42] V. Torra, *Hesitant fuzzy sets*, International Journal of Intelligent Systems, **25** (6) (2010), 529–539.

[43] J. Q. Wang, H. B. Li, *Multi-criteria decision-making method based on aggregation operators for intuitionistic linguistic fuzzy numbers*, Control and Decision, **25**(10) (2010), 1571–1574.

[44] X. K. Wang, H. G. Peng, J. Q. Wang, *Hesitant linguistic intuitionistic fuzzy sets and their application in multi-criteria decision-making problems*, International Journal for Uncertainty Quantification, **1** (2018), 1–10.

[45] J. Q. Wang, J. T. Wu, J. Wang, H. Y. Zhang, X. H. Chen, *Interval-valued hesitant fuzzy linguistic sets and their applications in multi-criteria decision-making problems*, Information Sciences, **288** (2014), 55–72.

[46] D. V. D. Weken, M. Nachtegael, E. E. Kerre, *Using similarity measures and homogeneity for the comparison of images*, Image and Vision Computing, **22**(9) (2004), 695–702.

[47] Z. S. Xu, *A method based on linguistic aggregation operators for group decision making with linguistic preference relations*, Information Sciences, **166**(1) (2004), 19–30.

[48] Z. S. Xu, *Deviation measures of linguistic preference relations in group decision making*, Omega, **33**(3) (2005), 249–254.

[49] Z. S. Xu, *Incomplete linguistic preference relations and their fusion*, Information Fusion, **7**(3) (2006), 331–337.

[50] Z. S. Xu, *Linguistic decision making: theory and methods*, Science Press, Springer, **7**(5) (2012), 479–492.

[51] Z. S. Xu, *Uncertain linguistic aggregation operators based approach to multiple attribute group decision making under uncertain linguistic environment*, Information Sciences, **168**(1) (2004), 171–184.

[52] Z. S. Xu, M. M. Xia, *On distance and correlation measures of hesitant fuzzy information*, International Journal of Intelligent Systems, **26** (2011), 410–425.

[53] X. Yang, Z. S. Xu, H. C. Liao, *Correlation coefficients of hesitant multiplicative sets and their applications in decision making and clustering analysis*, Applied Soft Computing, **61** (2017), 935–946.

[54] J. Ye, *Multicriteria fuzzy decision-making method using entropy weights-based correlation coefficients of interval-valued intuitionistic fuzzy sets*, Applied Mathematical Modelling, **34**(12) (2010), 3864–3870.

[55] C. H. Yu, *Correlation of fuzzy numbers*, Fuzzy Sets and Systems, **55**(3) (1993), 303–307.

[56] S. M. Yu, J. Wang, J. Q. Wang, et al., *A multi-criteria decision-making model for hotel selection with linguistic distribution assessments*, Applied Soft Computing, **61** (2017), 741–755.

[57] L. A. Zadeh, *Fuzzy sets*, Information Control, **8**(3) (1965), 338–353.

[58] L. A. Zadeh, *The concept of a linguistic variable and its application to approximate reasoning*, Information Sciences, **8**(4) (1975), 301–357.

[59] Q. S. Zhang, S. Y. Jiang, *A note on information entropy measure for vague sets and its applications*, Information Sciences, **178**(21) (2008), 4184–4191.

[60] X. Y. Zhang, H. Y. Zhang, J. Q. Wang, *Discussing incomplete 2-tuple fuzzy linguistic preference relations in multi-granular linguistic MCGDM with unknown weight information*, Soft Computing, **1** (2017), 1–18.